Niveau(x) d'enseignement : 0 Numéro EDUQ : 9570105
Centre d'intérêt : 841

Problèmes didactiques de l'évaluation du savoir mathématique. -- Montréal : Université du Québec à Montréal, Centre interdisciplinaire de recherche sur l'apprentissage et le développement en éducation, 1992.
1 h 15 min : son., coul.

Conférencier : Antoine Bodin. Également disponible en version imprimée, dans la coll. Séminaire sur la représentation, no 85. Offert sur vidéocassette.

Vedette(s)-secondaire(s) : Bodin, Antoine • Université du Québec à Montréal. Centre interdisciplinaire de recherche sur l'apprentissage et le développement en éducation

Descripteur(s) : Mathématiques • Évaluation de l'apprentissage • Évaluation formative • Évaluation sommative • Évaluation diagnostique

Le conférencier traite de la place que prennent les faits d'évaluation dans l'enseignement des mathématiques. Il passe en revue les différents types d'évaluation : interne, externe, formative, sommative, diagnostique, etc.

CENTRE INTERDISCIPLINAIRE DE RECHERCHE SUR L'APPRENTISSAGE ET LE DEVELOPPEMENT EN EDUCATION

clrade

Université du Québec à Montréal

95 70105

PROBLÈMES DIDACTIQUES DE L'ÉVALUATION DU SAVOIR MATHÉMATIQUE

par Antoine Bodin
Faculté des sciences et des techniques
IREM de Besançon

Séminaire sur la Représentation

3 novembre 1992 - no 88

CIRADE, UQAM, avril 1993

ISBN 2-921558-10-6
Dépôt légal - Bibliothèque nationale du Québec, 1995

Pour la préparation de ce texte, je dois beaucoup à la réflexion et aux échanges qui ont été menés au cours des années 1988-1990 au sein du groupe thématique «Les instruments de gestion de l'enseignement: objectifs - curriculums - évaluation» du Groupement de recherche didactique et acquisition des connaissances scientifiques du CNRS.

Je suis aussi largement redevable à la Direction des Lycées et Collèges du Ministère de l'éducation nationale, pour la disponibilité qu'elle m'a accordée me permettant de travailler dans le cadre de ce groupement de recherche.

Enfin je dois beaucoup au travail des nombreux enseignants et militants de l'APMEP[1] qui, depuis 1987, ont participé au développement de l'observatoire EVAPM[2].

Toutefois, dans l'état actuel d'avancement des travaux relatifs au domaine étudié, les propos tenus dans cet article n'engagent que leur auteur.

La plupart des questions d'évaluation utilisées au cours de mon exposé peuvent illustrer des points différents. Pour la présentation écrite, il m'a semblé préférable de les regrouper dans la dernière partie.

---

1 APMEP: Association des professeurs de mathématiques de l'enseignement public.

2 EVAPM: Dispositif développé dans le cadre de l'APMEP avec l'aide de plusieurs IREM et de l'INRP pour suivre l'évolution de l'enseignement des mathématiques; en particulier l'évolution consécutive à des modifications de programmes. L'observatoire utilise notamment des enquêtes à grande échelle.
La plupart des questions d'évaluation présentées dans ce document proviennent de l'observatoire EVAPM.

INTRODUCTION

La recherche en didactique des mathématiques, malgré l'importance de ses récents développements a, jusqu'à présent, produit peu de travaux explicites sur les phénomènes liés à l'évaluation.

Une des raisons que l'on peut invoquer pour expliquer cette mise entre parenthèses est sans doute que les faits d'évaluation se constituent en un réseau quasiment indépendant du système didactique et qu'ils semblent ainsi appartenir à un autre champ, ou relever d'une autre problématique.

La relative absence de travaux spécifiques ne signifie cependant pas que les didacticiens n'aient pas, depuis longtemps, été conscients de l'importance des phénomènes d'évaluation. Pour Yves Chevallard, par exemple, (Chevallard, Y., Feldmann, S., 1986) lorsque« *le didacticien tente de pénétrer dans l'histoire d'une classe, il doit se rendre à l'évidence: les faits d'évaluation qu'il peut alors y observer ne sont pas simplement un existant contingent, un mal nécessaire que l'on pourrait ignorer, mais bien l'un des aspects déterminants du processus didactique - qui règle et régule tout à la fois les comportements de l'enseignant comme l'apprentissage des élèves. Bref, quiconque pénètre un peu longuement dans la vie d'une classe ne peut longtemps ignorer la «tyrannie» du processus d'évaluation.*»

Toutefois, la recherche en didactique des mathématiques, même lorsqu'elle n'est pas dirigée directement vers les phénomènes liés à l'évaluation, conduit à des observations et à des savoirs qui modifient profondément les conceptions que l'on peut se faire de l'évaluation.

Le présent exposé se propose de tenter une mise en perspective des faits d'évaluation, de la façon dont ces faits interrogent la didactique des mathématiques, des réponses que la didactique des mathématiques peut d'ores et déjà apporter à certaines de ces questions et des orientations qui se dégagent.

Les recherches spécifiques sont encore trop embryonnaires et trop disparates pour qu'il soit possible de proposer une synthèse claire. L'exposé, même s'il cherche à structurer l'analyse entreprise, n'échappera pas à un certain pointillisme. Il s'attachera surtout à dégager des pistes et à ouvrir des perspectives.

Compte tenu de l'aspect systémique de l'organisation des faits d'évaluation, il ne nous semble pas possible de nous centrer d'emblée sur la vie de la classe et sur la relation didactique qui y prend place. Dans un premier temps, au contraire, il nous paraît nécessaire de prendre en compte les fonctions très diverses assurées par les dispositifs qui se réclament de l'évaluation.

D'une part, les réflexions qui suivent prennent leur origine dans la familiarité que l'auteur entretient avec un système éducatif particulier: le système éducatif français et plus particulièrement, son système d'enseignement secondaire des mathématiques. Toutefois, ses interventions dans d'autres pays comme son implication personnelle dans la Troisième Étude Internationale sur l'Enseignement des Mathématiques et des Sciences (TIMSS de l'IEA) devrait permettre de se situer dans un cadre plus large.

En ce qui concerne les faits d'évaluation, il est clair qu'ils sont très variables d'un pays à l'autre, et très dépendant des coutumes et des cultures des milieux ou des pays concernés. Malgré les précautions prises, il est certain que les propos qui suivent, risquent de présenter quelques biais culturels et que des ajustements pourront être nécessaires pour les adapter aux conditions locales particulières.

Derrière la petite phrase: *problèmes didactiques de l'évaluation du savoir mathématique* se cachent autant de pièges que de mots. L'acception donnée à chacun de ces mots est loin d'être univoque, même à l'intérieur d'un système particulier. Les choses se compliquent lorsque l'on cherche à généraliser ou que l'occasion nous est donnée de communiquer nos réflexions à des personnes participant d'un autre système, d'une autre discipline que les mathématiques, d'un autre pays que la France, d'autres niveaux de scolarité. Nous aurons l'occasion de voir que même les objets de savoir ne sont pas stables lorsque l'on passe certaines de ces frontières!

Une définition des termes utilisés nous a semblé indispensable. Mais avant, nous allons commencer par effectuer un premier repérage des faits d'évaluation.

Place de l'évaluation dans le système d'enseignement des mathématiques

Sous des pressions diverses, la notion d'évaluation des pratiques qui l'accompagnent ont, au cours des trente dernières années, fini par investir totalement le champ de l'éducation. De ce fait, l'observateur du système d'enseignement des mathématiques se trouve

aujourd'hui confronté à un grand nombre de faits qui lui sont présentés comme relevant de l'évaluation.

Ces faits concernent aussi bien l'action quotidienne de l'enseignant (préparer des devoirs, corriger, mettre des notes,...) que la planification de l'enseignement, le fonctionnement de l'établissement, etc. Ils concernent les comptes que la société demande à son système éducatif.

La figure 1 essaie de mettre en relief quelques-uns de ces faits en les situant par rapport aux divers constituants du système. Il est en effet de plus en plus manifeste que les diverses actions d'évaluation forment une sorte de réseau dont les éléments interviennent les uns sur les autres. Ainsi, il sera difficile de comprendre l'évaluation qui se fait à un moment donné dans une classe donnée, si l'on ne sait rien de l'évaluation institutionnelle qui l'a précédée ou d'une évaluation en cours concernant le fonctionnement de l'établissement.

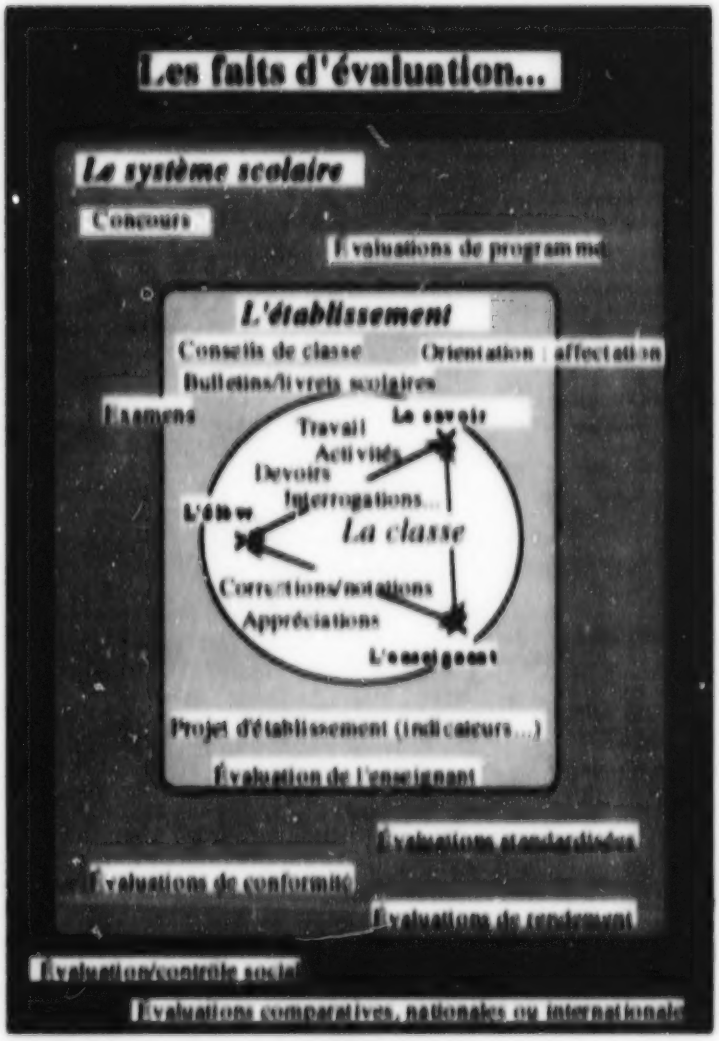


Figure 1

En France, par exemple, étudier le fonctionnement de l'évaluation dans une classe de lycée, sans prendre en compte l'existence du baccalauréat, ne serait sans doute pas raisonnable. À la limite, on peut même se demander s'il est possible de s'intéresser à l'évaluation des élèves sans s'intéresser aux représentations que les enseignants ont de l'évaluation, donc sur ce qui détermine ces représentations et en particulier, sur la façon dont ils sont eux-mêmes évalués.

Dans cet exposé, nous nous limiterons cependant aux faits d'évaluation qui mettent directement en jeu le savoir mathématique.

Qu'il s'agisse de l'évaluation que les enseignants font directement de leurs élèves, de l'évaluation institutionnelle, ou encore de celle, indirecte, faite par le corps social et les médias, il y a toujours intervention d'un **jugement** sur la qualité des connaissances acquises par les élèves, ou par les étudiants. Comment se construisent ces jugements? Quel est leur degré de consistance et de pertinence? Par rapport à quelles conceptions? Voilà qui ne peut qu'intéresser la recherche en didactique!

Dans les discours comme dans les pratiques relevant de l'évaluation, on observe une oscillation entre divers pôles. Ainsi l'évaluation intervient-elle tantôt avec le propos de juger, de classer, ou de certifier, tantôt avec celui de comprendre les processus et d'aider l'apprenant. Tantôt se propose-t-elle d'être un outil de communication, permettant à l'élève de mieux comprendre les enjeux de ses apprentissages, tantôt elle est utilisée comme moyen de lui imposer des comportements ou des objectifs.

D'une façon générale et ne recouvrant que partiellement les distinctions ci-dessus, l'évaluation oscille entre le désir de mesurer à tout prix et la volonté d'expliquer, de donner du sens aux observations qu'elle permet de faire.

PROBLÉMATIQUES EN PRÉSENCE

Avant d'aller plus loin, il semble indispensable de préciser la signification que nous donnons à quelques-uns des termes utilisés dans cet exposé. Cela d'autant plus que certains de ces termes ont pris un sens particulier, en France et dans quelques autres pays, au moins dans certains milieux.

Une didactique consiste en une actualisation d'un projet social de faire acquérir des connaissances à autrui, éventuellement de transmettre des valeurs culturelles.

Ainsi pourra-t-on parler de la didactique mise en oeuvre par tel enseignant, proposée par tel manuel ou curriculum officiel. On pourra être plus précis et évoquer par exemple **une** didactique des décimaux ou de l'intégration.

La didactique est le champ qui se constitue à partir de la prise en compte des didactiques particulières.

> Selon Marc Legrand (exposé 1993 - non publié) *« il y a du didactique chaque fois qu'il y a une double intentionnalité: celle du professeur qui détient un savoir, et qui l'aménage pour le transmettre et celle de l'élève qui reconnaît l'intérêt de ce savoir et qui veut l'apprendre. »*

Cette définition paraît raisonnable et nous la faisons nôtre. Toutefois elle n'est pas de nature à nous faciliter la tâche. En effet, même si Marc Legrand ne dit pas qu'il donne là une condition **nécessaire** pour que l'on puisse reconnaître du didactique, il met implicitement en garde contre la tentation d'intégrer trop rapidement n'importe quoi dans ce champ.

En ce qui concerne les faits d'évaluation, notre hypothèse est que de façon directe ou indirecte, ils interviennent de façon prégnante, pour modeler la double intentionnalité en question. Il paraît cependant souhaitable de ne pas chercher à faire entrer de force tous les faits d'évaluation dans le champ du didactique, ce qui nous amène à proposer le schéma *figure 2*.

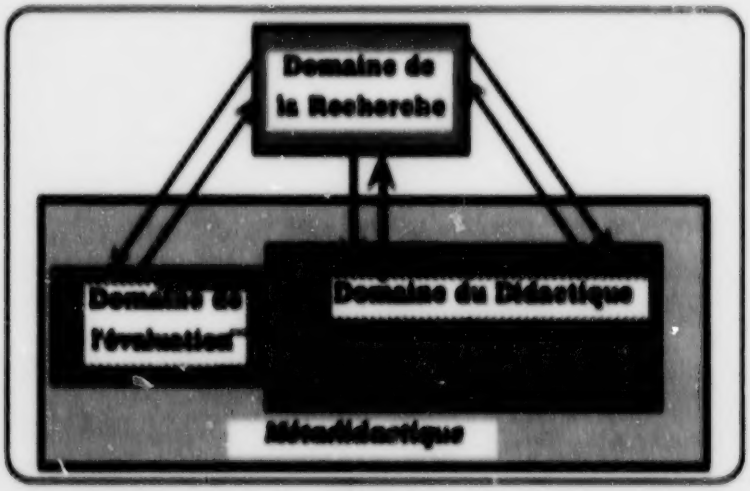


Figure 2

La recherche, dans un domaine quelconque, a pour objet de dégager des invariants, des lois. Elle conduit à généraliser et à produire des systèmes de repérage.

Pour ce qui nous concerne ici, la recherche peut porter sur une didactique particulière, elle peut aussi porter sur l'évaluation. On sait par exemple que la recherche en didactique des mathématiques a produit et utilisé un certain nombre de concepts sur lesquels elle s'appuie et qui permettent de mettre en évidence des régularités, sinon des lois: voir, par exemple, le

contrat didactique, la transposition didactique, l'(les) obstacle(s), le champ conceptuel... **On ne peut bien entendu pas confondre une** (ou la) recherche en didactique et une didactique particulière.

On parle donc de **recherches en didactique** pour parler de recherches qui portent sur **la didactique.**

Ces recherches ont eu besoin de s'appuyer sur un cadre théorique qui s'est peu à peu constitué et qu'elles ont aidé à constituer. C'est ce cadre théorique que nous nommons **la didactique**.

La didactique peut donc être définie comme une théorie, ou un ensemble de théories, modélisant la didactique.

Selon Marc Legrand, «*La didactique est le résultat d'un travail de théorisation qui vise à comprendre et à expliquer le réel de l'enseignement.*»

Une définition opératoire de l'évaluation

Il existe de nombreuses définitions de l'évaluation, souvent circonstancielles et incomplètes, voire contradictoires. Nous essayons ici de proposer **une** définition qui permette à la fois de rendre compte de la plupart des faits d'évaluation observés, de faciliter l'entrée dans le champ de la didactique et particulièrement, de la didactique des mathématiques, tout en permettant de communiquer avec les autres champs (sciences de l'éducation, psychologie, psychométrie,...).

L'**ÉVALUATION** consiste en l'ensemble des **procédures** et des **processus** dont la finalité première est de porter un jugement sur la «valeur» d'un «objet».
ÉVALUER suppose d'**organiser et d'étudier des situations** permettant de recueillir des informations qui, **après traitement,** soient susceptibles de révéler quelque chose de fiable et de substantiel sur la «valeur» d'un «objet».

L'**objet** de l'évaluation peut être de nature et d'ampleur très diverses. En nous restreignant à ce qui peut intéresser l'enseignement des mathématiques, cet objet peut être le «savoir» d'un élève dans un domaine donné, une situation d'enseignement, un programme d'enseignement, une question d'évaluation, une épreuve, un test, un examen, un manuel, un didacticiel, etc.

La **valeur** peut concerner la conformité (ou l'écart) d'un comportement à des attentes, la qualité d'une production (devoir), la qualité d'un apprentissage, la signification d'un comportement observé, la valeur formative ou informative d'une situation, etc. Elle peut être intrinsèque ou relative, de nature épistémologique, normative, critériée,...

Malgré la difficulté qu'elle apporte, l'idée de **valeur** ne peut pas être évacuée. Le schéma *figure 3* cherche à montrer que lorsque l'évaluation renonce à tenter de dire quelque chose sur la **valeur**, elle perd sa raison d'être: en effet, les pratiques résiduelles trouvent alors leur place dans l'organisation des situations didactiques sans qu'une référence quelconque à l'évaluation ne soit nécessaire (voir aussi *figure 2*).

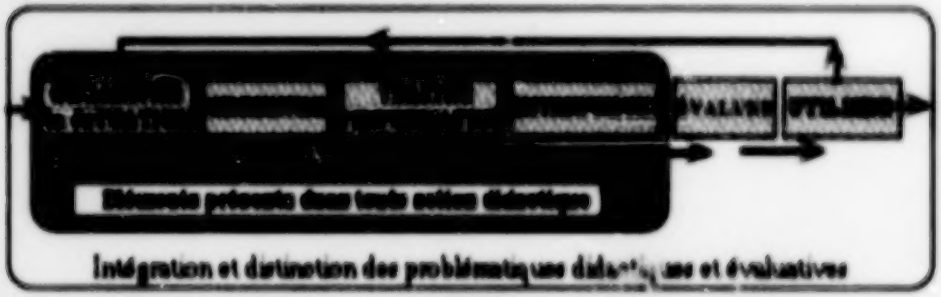


Figure 3

La définition proposée, avec l'insistance mise sur la valeur, nous semble de nature à favoriser l'exigence de rigueur et, autant que faire se peut, de «vérité», qui nous semble devoir caractériser les pratiques d'évaluation. On ne peut manquer d'évoquer à ce sujet la **fonction de vérification** de l'évaluation telle qu'elle est définie par Yves Chevallard (1989) et qui, selon l'auteur, serait en perte de vitesse, peut être sous les assauts, de certaines pratiques d'évaluation sur lesquelles nous aurons l'occasion de revenir.

Selon notre définition, qui s'éloigne ainsi, de celle proposée par Stufflebeam, **l'évaluation, *prépare et nourrit des décisions***: elle ne les impose pas. Bien entendu, la décision concerne l'*objet* évalué, mais elle reste extérieure à l'évaluation. Il peut s'agir d'une microdécision (concernant par exemple la poursuite ou la modification d'une séquence d'enseignement) aussi bien qu'une macrodécision (conseiller ou imposer un redoublement à un élève, changer de manuel, modifier un programme d'enseignement...).

FONCTIONS DE L'ÉVALUATION ET DIDACTIQUE

Il est devenu classique de distinguer différentes fonctions de l'évaluation et de leur associer des méthodologies différentes. Évaluation sommative, formative, diagnostique,..., démarche critériée, normative, etc.

Ces distinctions sont utiles lorsqu'il s'agit de décrire ou de spécifier des actions, de proposer une instrumentation. Elles ont cependant l'inconvénient de se diversifier à l'infini, sans qu'il soit toujours facile de voir ce qui les distingue et ce qu'elles ont en commun.

D'autre part, ces distinctions concernent essentiellement les procédures mises en œuvre et nous souhaitons ici nous intéresser plus particulièrement aux processus: comment et pourquoi l'évaluation fonctionne, comment elle interagit avec les apprentissages, etc. La tentation est alors d'intégrer l'étude des faits d'évaluation dans l'étude d'un processus plus général qui serait celui de la régulation du système d'enseignement.

Un concept intégrateur, le concept de régulation

On peut définir la régulation comme l'ensemble des processus et des procédures assurant l'équilibre d'un système. Ainsi définie, la notion de régulation d'un système didactique ne contient pas l'idée d'asservissement aux objectifs poursuivis (ce qui est toutefois le cas pour la régulation d'un système finalisé). Dans le cas général cependant, la régulation a lieu, mais elle peut éventuellement éloigner du but poursuivi.

Remarquons déjà qu'il n'est pas possible de considérer d'emblée toute didactique, au sens défini ci-dessus, comme un système finalisé dont la finalité réelle correspondrait à la finalité affichée.

L'évaluation, quelles que soient les fonctions affichées, contribue de façon évidente à la régulation du système didactique. Cependant, la régulation ne se réduit pas à l'évaluation et s'intéresser à la régulation serait s'intéresser à un objet beaucoup plus «gros» que celui qui pourrait être constitué de l'ensemble des faits d'évaluation.

Pour l'instant au moins, il nous semble donc nécessaire de continuer à parler d'évaluation et même des différentes fonctions affichées, cela au moins pour ne perdre de vue l'intentionnalité des acteurs et son éventuelle congruence ou contradiction avec les effets observés.

Évaluer pour juger ou évaluer pour aider?

Remarquons qu'à propos d'évaluation, l'Encyclopédia Universalis renvoie directement et uniquement à l'article **Jugement**. Qu'il s'agisse d'évaluer pour désigner les étudiants qui seront reçus à tel examen, pour **certifier** la qualité d'une formation, ou pour se prononcer sur la compétence professionnelle d'un enseignant, c'est bien, en fin de compte, de jugement qu'il s'agit.

Dans ce contexte le terme d'**évaluation sommative** convient bien. Il s'agit alors d'établir ou d'inférer un portrait suffisamment complet pour que la décision prise ne dépende pas du tableau particulier produit.

En France, les recherches docimologiques, dont la tradition remonte à Pieron (1920), portent le plus souvent sur l'examen du baccalauréat et donc sur l'évaluation sommative. Ces recherches ne se déroulent habituellement pas dans le cadre didactique et ne prennent pas en compte la spécificité des savoirs en question. Nous n'en parlerons donc pas directement ici.

La plupart des recherches et des ouvrages récents, relatifs à l'évaluation, portent sur l'évaluation **formative**, c'est-à-dire sur les pratiques d'évaluation dont l'objectif annoncé est l'amélioration de la formation individuelle ou collective (on parlera ainsi d'évaluation formative des programmes, comme le faisait Scriven dès 1967).

L'évaluation formative se caractérise par une intention (déclarée), par des pratiques et par des effets. Voici, par exemple, deux citations qui s'apparentent à des définitions potentielles ou virtuelles: Pour Jean Cardinet (1978): l'évaluation formative «***a pour but de guider l'élève dans son travail scolaire. Elle cherche à situer ses difficultés pour l'aider à découvrir des procédures qui lui permettent de progresser dans son apprentissage.***»

Pour Philippe Perrenoud (1991): «*Est formative toute évaluation* ***qui aide*** *l'élève à apprendre et à se développer, autrement dit,* ***qui participe*** *à la régulation des apprentissages et du développement dans le sens du projet éducatif.*»

On remarquera que la première définition porte sur l'intention alors que la seconde porte sur les effets de l'action. Dans la définition de Perrenoud, on peut d'ailleurs remplacer le mot **évaluation** par *action, situation, démarche...* et même par *didactique*.

On touche là une question d'importance: le discours pédagogique ne permet pas toujours de distinguer entre instruments, démarches et méthodes... concepts impliquant l'atteinte de certains buts et de l'effectivité de ces mêmes objets ou actions. Par exemple, on peut lire une phrase telle que «*on apprend aux élèves à ...*» alors que l'on ne parle que de la stratégie utilisée et que, peut-être, soumis à cette stratégie, certains élèves n'apprennent rien. Pour le verbe **apprendre**, on a cependant l'avantage de pouvoir distinguer **enseigner** et **apprendre**, mais rien de tel pour **évaluation formative**: tantôt désir, tantôt «réalité». Rien de tel non plus pour **formation** qui peut désigner aussi bien une action qu'un produit.

Ainsi, il faudrait dire «*une évaluation formative (démarche) réellement formative (effet) est une évaluation qui concourt à ce que la formation donnée (par le biais d'une didactique) conduise à une formation réussie (dans un sens à définir) pour la cible de l'évaluation*». Mais cela devient bien confus!

Une façon de simplifier consiste à confondre des notions que la raison commande de distinguer et de pratiquer une sorte de syncrétisme où tout et n'importe quoi peut être dit. Une autre consiste bien sûr à assumer la distinction, ce que les auteurs cités plus haut essayent de faire pour leur part, et ce que je vais essayer de continuer à faire ici.

Ainsi une démarche d'évaluation, formative dans ses intentions, peut être sommative de fait et inhibitrice de formation dans ses effets, tandis qu'une démarche d'évaluation, sommative dans ses intentions, peut avoir des effets formatifs tout à fait positifs.

Évaluation et contrat didactique

La notion de **contrat didactique**, telle qu'elle est définie et utilisée par Guy Brousseau (1986), est ici essentielle. Insistons sur le fait que, dans la modélisation adoptée, l'enseignant n'a pas le pouvoir de décider de ce qui entre ou n'entre pas dans le contrat, de ce qui le modèle. Un **contrat pédagogique** explicite peut cohabiter avec le contrat didactique, éventuellement l'influencer; les deux notions ne peuvent en aucun cas être confondues.

Nous renvoyons ici aux textes [illegible]teurs de Guy Brousseau ainsi qu'aux analyses qui ont été faites en termes de contrat de phénomènes tels que celui de «l'âge du capitaine» (Chevallard, 1988) et dont les rapports avec la question de l'évaluation sont évidents. La suite du texte nous donnera l'occasion de rencontrer plus loin quelques manifestations du contrat dans des situations d'évaluation.

Nous faisons l'hypothèse que l'évaluation apparemment la plus éloignée de la formation, l'évaluation sommative, la plus administrative qu'il soit possible de concevoir, imprégne de sa marque le contrat didactique (de façon éventuellement préventive). Par contre, la part de l'évaluation qui est intégrée aux situations d'apprentissage, qui joue un rôle positif dans les apprentissages et qui se trouve totalement intégrée au contrat didactique ne peut plus guère être pensée de façon indépendante de la didactique.

Nous renvoyons à nouveau ici aux schémas des *figure 2* et *figure 3* qui, d'une certaine façon, illustrent ce qui précède.

L'évolution des conceptions et des connaissances concernant l'apprentissage conduit de plus en plus souvent à intégrer aux situations d'apprentissages relevant, par exemple, de conceptions constructivistes, des éléments qui, dans une conception béhavioriste auraient été considérés comme extérieurs et relevant strictement de l'évaluation.

En mathématiques, comme dans le cadre de la classe, il n'est plus guère concevable de concevoir l'enseignement et l'évaluation comme deux actions parallèles. La première, l'enseignement, consistant à déverser des connaissances plus ou moins formelles, tandis que l'autre, s'attacherait à contrôler que les élèves développent alors des connaissances opératoires, comme en écho.

Si l'on prend en compte, par exemple, la position de Vergnaud (1981), selon laquelle la résolution de problème est «source et critère de savoir», l'élève devra être confronté à des problèmes aussi bien si l'on souhaite lui faire acquérir des connaissances que si l'on souhaite contrôler l'existence de ces connaissances (ou leur développement). Toujours en mathématiques, une partie au moins de l'évaluation (telle que nous l'avons définie) est strictement interne. La nécessité de vérifier les résultats obtenus et celle de se prononcer sur la valeur des procédures utilisées, font partie de l'activité normale du mathématicien, comme de l'utilisateur des mathématiques. Ces nécessités doivent s'inscrire dans la définition des compétences du domaine mathématique et donc dans les tâches proposées aux élèves. Cet aspect de la question a été largement développé et utilisé par Guy Brousseau (Brousseau, 1986,...), sous le nom de **dialectique de la validation** et repris de façon particulièrement éclairante par Claire Margolinas (1993).

Si les conceptions plus ou moins constructivistes des apprentissages modifient la place relative de l'enseignement et de l'évaluation, que dire alors de la prise en compte éventuelle, par les didactiques et par l'évaluation, du concept de **zone proximale de développement** de Vygotski?

Rappelons que (Vygotski, 1935 - repris 1998):

> *«La zone proximale de développement peut être considérée comme limitée d'un côté par ce que l'apprenant est capable de faire, le savoir qu'il peut manifester de façon indépendante et autonome et de l'autre côté par ce dont il serait capable s'il était guidé par l'enseignant, s'il travaillait en collaboration avec des pairs, ou s'il disposait d'aides adaptées (instrumentation...).»*

Au niveau des recherches comme au niveau des pratiques, la référence aux travaux et aux idées de Vygotski et en particulier à la notion évoquée ci-dessus, conduit à privilégier une **évaluation dynamique** portant sur les acquisitions immédiatement possibles et non plus seulement sur les acquisitions déjà effectuées. Cela, non seulement pour organiser les apprentissages, mais aussi pour certifier des compétences.

Zone actuelle de développement (ZAD)
et Zone proximale de développement (ZPD)
pour un domaine donné
ZPD de l'élève A
ZPD de l'élève B
ZAD des élèves A et B
ZPD de deux sujets ayant même ZAD

Figure 4

Pour employer les termes utilisés par Vygotski, il s'agit de passer d'une évaluation rétrospective, à une évaluation prospective.

Toutefois, alors qu'il est clair que cette référence peut être utilisée, sans doute avec profit, pour l'organisation des situations d'enseignement. Les questions qu'elle est susceptible d'apporter ou de résoudre, dans le domaine de l'évaluation, n'apparaissent pas encore très clairement.

L'un des pièges, à propos d'évaluation est, lorsqu'on se réfère à Vygotski contrairement à ce qui est proposé plus haut, la tentation de faire sortir des éléments le champ strict du didactique alors qu'ils s'y trouvaient fort bien, pour les introduire dans le champ de l'évaluation et donc, d'accentuer ce que nous avons appelé la pression évaluative. Cependant, une opérationnalisation réussie des notions vygotskiennes pourrait conduire à une définition dynamique des compétences.

Les recherches en didactique des mathématiques, se référant de plus en plus souvent à Vygotski, on peut cependant espérer qu'elles contribueront à éclaircir la notion de compétence (au sens large de virtualité) et qu'elles pourront contribuer à améliorer ce qui reste de la partie sans doute la moins bien maîtrisée de l'évaluation, je veux dire l'évaluation certificative, avec ce qu'elle comporte d'évaluation nécessairement sommative.

Intégration des fonctions et des problématiques

Nous avons vu, à plusieurs reprises que des pans entiers de la problématique évaluative se dissolvaient dans la didactique. Toutefois, nous pensons que cette intégration ne doit pas être forcée et que la spécificité de certains faits d'évaluation doit continuer à être affirmée et assurée.

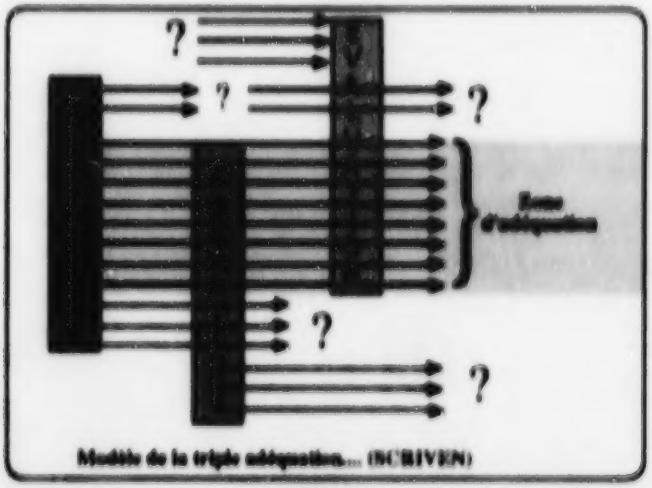
En contrepartie, nous pensons que les interventions correspondantes demandent à être justifiées et, en particulier, devront faire la preuve de leur utilité et, autant que possible, montrer qu'elles ne contrecarrent pas l'action didactique.

Certes, les informations que l'on a besoin de recueillir lors d'une évaluation du «rendement» d'un système de formation, ou celles que l'on souhaite recueillir lors d'une évaluation certificative (examen externe), ne sont pas nécessairement de même type et de même ampleur que celles que l'on cherche à recueillir lors de l'évaluation «formative» pratiquée au quotidien de la classe. Cependant, il n'y a pas de raison pour que les informations obtenues dans les premiers cas ne puissent être utilisées à des fins formatives.

Il n'y a pas non plus de raison pour que les tâches auxquelles ont soumises les élèves dans les premiers cas ne puissent être reconnues comme pertinentes par les spécialistes de la discipline et par les enseignants. Il y a encore moins de raison pour que les enseignants soient incités à pervertir leur action pour inciter les élèves à la maîtrise de tâches auxquelles ils n'accordent que peu de pertinence.

Cela, se vérifie à tout le moins dans le cas où il est possible d'accepter l'idée que l'épistémologie des enseignants n'est pas trop défaillante (dans le cas contraire un effort de formation sera nécessaire). C'est le cas lorsque l'évaluation d'un système éducatif donné est faite à partir de batteries de questions à choix multiples conçues dans, et pour, un autre système. C'est encore plus le cas lorsque l'analyse de contenu la plus élémentaire met en évidence que seuls sont pris en compte des connaissances procédurales qui ne recouvrent qu'une partie des savoirs que le système se propose de développer. Lorsque, enfin, on sait que la «mesure» effectuée à partir de ces batteries est faiblement corrélée avec ce que l'on obtiendrait avec des épreuves mettant en jeu d'autres types de savoirs, par ailleurs jugés essentiels.

L'effet en retour sur les représentations des enseignants, sur leurs pratiques et, finalement sur la constitution du contrat didactique fait qu'il n'est plus possible de s'en désintéresser. La recherche en didactique doit aussi s'intéresser aux évaluations qui se déclarent étrangère à l'action didactique proprement dite.



Modèle de la triple adéquation... (SCRIVEN)

Figure 8

Une question d'évaluation utilisée dans un contexte donné n'est acceptable, selon nous, que si elle garde de l'intérêt dans tout autre contexte. Elle peut être étrangère à un domaine particulier ou moins pertinente qu'une autre, elle peut être d'utilisation moins bien adaptée à la formation que l'on veut donner ou à l'information que l'on souhaite recueillir, elle peut ne pas avoir les propriétés psychométriques souhaitables dans une situation particulière,... mais elle ne doit pas pouvoir être rejetée comme définitivement étrangère aux résultats attendus d'une formation en mathématiques. En particulier, c'est une question qui continuera à apporter une information fiable et utile sur des connaissances reconnues utiles.

Le schéma de la *figure 5* reprend la vieille idée de la triple adéquation objectifs - formation - évaluation de Scriven (1967) nous paraît bien illustrer cette demande d'une évaluation qui reste cohérente dans ses diverses manifestations.

Il ne s'agit plus seulement d'intégration d'une partie de l'évaluation formative dans la didactique, il s'agit de faciliter le glissement suggéré par le schéma ci-dessous.

Évaluation sommative ⟶ Évaluation formative ⟶ Didactiques

En fait, les didactiques ne peuvent que chercher à rejeter les pratiques d'évaluation qui contrarient leur projet et assimiler celles qui les facilitent. Dans certains cas, il faudrait même parler d'accommodation car, comme nous l'avons vu, ce sont les didactiques elles-mêmes qui se transforment pour intégrer les pratiques d'évaluation.

Signalons que la question des relations entre les pratiques d'évaluation formative et les pratiques didactiques a été soulevée ces derniers temps, de façon explicite, par plusieurs auteurs: en particulier, par Daniel Bain (1988), à propos de la didactique du français et plus généralement par Philippe Perrenoud (1991). Cette question me semble d'autre part être présente de façon implicite dans de nombreux autres travaux, en particulier ceux qui cherchent à prolonger ou à renouveler les théories concernant l'évaluation formative.

Évaluer pour communiquer ou évaluer pour conditionner?

S'intéresser à l'évaluation comme productrice d'information utile sur le savoir des élèves sur son existence, sa consistance, son devenir, est-ce faire fausse route?

Si l'on en croit Yves Chevallard (1986), «*la note assignée par le correcteur n'est pas mesure, mais message. Ce message intervient dans une négociation, ou une transaction, qui signe un rapport de forces entre l'enseignant et les enseignés, à propos du savoir enseigné.*»

Une autre façon de mettre en doute l'intérêt d'une recherche de consistance de l'évaluation en tant que productrice d'information concerne la relation, que de façon évidente, l'évaluation entretient avec l'**institutionnalisation**. Citons par exemple Johsua et Dupin (1993): «*L'évaluation redouble... à sa manière la fonction de l'institutionnalisation. Elle confirme, dans le cadre d'un contrat éventuellement spécifique à une classe, ce qui doit être considéré comme important, et ce qui est secondaire, ce qu'il est décisif de savoir faire et ce qui est accessoire.*»

Il est à noter que la remarque faite par Chevallard ne semble concerner que la relation didactique proprement dite et que de plus, elle ne concerne que la note, ce qui ne correspond qu'à une façon de communiquer le jugement consécutif une évaluation. La seconde remarque est plus générale.

On peut cependant admettre l'idée que toute évaluation est le prétexte d'une négociation plus ou moins importante. Même au niveau des examens ou des évaluations du système éducatif, la publicité (et parfois la non publicité!) faite aux dispositifs et aux résultats peut être assimilé à une négociation.

L'importance de la communication dans l'évaluation est soulignée depuis longtemps. On ne parlait pas encore d'évaluation en France dans la première moitié de ce siècle, mais on organisait de cérémonies imposantes pour communiquer aux élèves les résultats de leurs efforts: je veux bien sûr parler des distributions des prix.

De nombreux travaux ont progressivement mis en évidence les effets positifs sur les apprentissages de la communication préalable des objectifs, puis celle de la communication des critères de l'évaluation (voir par exemple Bonniol, 1981). Des travaux plus récents ont montré l'efficacité de la communication directe des produits attendus ainsi que des procédures supposées être mises en oeuvre.

Écoutons par exemple Michèle Genthon (1991), parlant de l'élève en apprentissage: «*non seulement il faut lui communiquer les critères qui permettent d'évaluer les produits, les résultats, les performances, mais ceux qui portent sur les procédures permettant*

d'obtenir ces résultats. En amont encore, il convient d'expliciter les critères relatifs aux stratégies à développer, aux opérations à mener, constitutives de ces stratégies».

Cette conception de la communication liée à l'évaluation associée à la prise en compte de la notion de «base d'orientation de l'action» de Galperine (1966) a conduit au moins en France à une instrumentation particulière d'évaluation pédagogique connue sous le nom **d'évaluation formatrice**.

Dans le cas de cette démarche, il y a indubitablement intégration d'une partie de l'évaluation à la didactique utilisée, modification profonde de cette didactique et modification du contrat. Les observations faites concluent à un effet important sur les compétences développées, telles du moins qu'elles sont évaluées ensuite. Nous n'avons pas de raison d'en douter.

Pour ce qui est de l'utilisation d'une telle démarche dans le cadre de l'enseignement des mathématiques, on peut cependant s'interroger sur le risque de perversion du sens des connaissances ainsi acquises. En particulier, on peut en effet se demander ce que deviennent les phases de validation internes. Le risque n'existerait peut-être pas dans le cas où l'adéquation objectifs - formation - examen (cf. *figure 5*) serait parfaite. Dans le cas contraire on peut craindre une adaptation à l'examen qui contribue un peu plus à l'évanouissement du sens des connaissances acquises contre lequel Guy Brousseau met en garde depuis longtemps. Citons encore Johsua et Dupin (1993):

> *«Dans sa volonté de réduction de l'ambiguïté née du processus de passage de la gestion professorale du rapport au savoir à celle de l'élève, le système d'enseignement penche de fait vers l'évanouissement du sens de ces savoirs.»*

À propos des inquiétudes manifestées par Guy Brousseau par rapport à l'évaluation en général, et pas seulement pour ses aspects liés à la communication, il nous semble utile de citer un peu plus longuement cet auteur (Brousseau, 1989):

> *«Beaucoup d'enfants et ensuite de professeurs n'acquièrent qu'une fausse pratique du savoir... L'évaluation seule ne permet pas de corriger les phénomènes rapportés plus haut et parfois elle les accentue...*
>
> *1 - L'évaluation permet à certains maîtres de pallier en partie aux insuffisances les plus criantes de leur enseignement.*
> *2 - Elle le conduit ensuite à un diversification et à une démultiplication des objectifs intermédiaires, donc à recourir à des situations formées visant des apprentissages à court terme, avec pour corollaire l'effet de rationalisation.*
> *3 - Enfin, tôt ou tard l'évaluation se limite à certains objectifs convenus. Les maîtres les visent par des apprentissages d'algorithmes et de savoirs non*

fonctionnants. L'économie conduit à céder sur les exigences relatives à la compréhension...

Il n'est pas encore possible de distinguer, à l'aide des évaluations classiques, des connaissances acquises par une suite organisée d'assimilations ou même de conditionnement de celles qui sont acquises par une genèse authentique des concepts.»

D'une autre façon, Claire Margolinas met en garde, de façon indirecte, contre les procédures qui risquent de conduire à l'illusion du savoir, ou à un savoir qui se couperait de ses racines et de ses significations profondes (Margolinas, 1993):

«Si l'élève cherche dans les savoirs mathématiques que veut lui transmettre le maître une collection de recettes erratiques, s'il n'est pas passé de l'empirisme au rationalisme en ce qui concerne son rapport personnel aux mathématiques enseignées, le maître ne peut l'y faire rentrer «de force»... L'entrée dans une problématique apodictique représente à la fois les étapes initiales et finales d'un apprentissage de mathématiques. Ce lien crucial entre mathématiques et vérité apodictique place le point de vue de la validation au centre des problèmes d'enseignement des mathématiques.»

La prise en compte systématique des aspects communicationnels de l'évaluation peut être poussée à des limites à côté desquelles les pratiques traditionnelles d'évaluation paraîtront bien innocentes. Dans une thèse récente soutenue à Ottawa, Charles Fournier (1993) met clairement en évidence l'avantage qu'il peut y avoir à entraîner les étudiants à **«anticiper»** l'examen. L'idée n'est pas nouvelle et de nombreux pédagogues ont suggéré aux étudiants de «se mettre en scène» et de répéter l'examen, particulièrement dans le cas de l'oral. Ce qui est sans doute nouveau, c'est la tentation de systématisation. Dans l'expérimentation, les étudiants sont en effet amenés à prévoir les questions qui leur seront posées mais aussi à faire un pronostic sur leurs chances de les traiter correctement et cela, avant même d'essayer de les traiter.

La recherche de Charles Fournier n'est pas en cause. Il s'agit d'une recherche rigoureuse qui, en tant que recherche, nous apprend des choses utiles sur les processus d'évaluation et sur les conduites face à l'évaluation. Il est même vraisemblable que des applications systématiques pourront être utiles pour certaines formations.

Mais nous parlons ici de mathématiques et nous avons déjà essayé de montrer que, dans ce domaine, l'entrée par l'évaluation, par l'examen et par l'adaptation à l'examen, ne pouvait conduire qu'à une perversion épistémologique. De plus, C. Fournier construit une partie de son argumentation sur le schéma de la *figure 6*. En particulier, le «domaine utilisé par l'enseignant» serait ainsi une partie du «domaine du spécialiste».

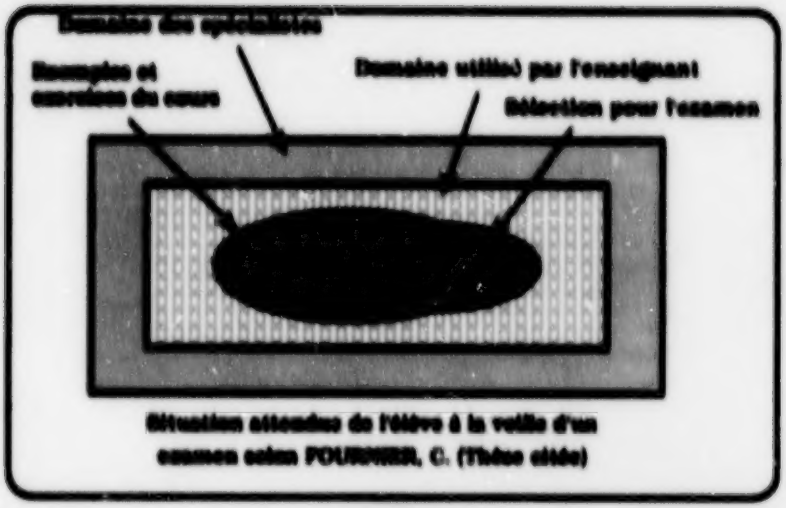

Domaine des spécialistes
Exemples et exercices du cours
Domaine utilisé par l'enseignant
Sélection pour l'examen
Situation attendue de l'élève à la veille d'un examen selon FOURNIER, G. (Thèse citée)

Figure 6

Dans le cas des mathématiques, «le domaine du spécialiste» renvoie bien sûr au **savoir savant** et l'ensemble du schéma nous renvoie à la **transposition didactique**.

En mathématiques, comme Yves Chevallard (1985) l'a abondamment montré, le savoir officiel (scolaire) ne peut être considéré comme simplement inclus dans le savoir savant. De même, le savoir enseigné ne peut être identifié au savoir officiel pas plus que les questions de l'examen ne peuvent être considérées de façon systématique comme des éléments du «domaine utilisé par l'enseignant» (savoir enseigné?).

À moins de parvenir à la parfaite adéquation déjà évoquée, la situation de l'étudiant au moment de l'examen a plus de chance de correspondre au schéma ci-contre (*figure 7*).

Des questions de pertinence se posent alors: tel élément de savoir est-il utile? Pourquoi?

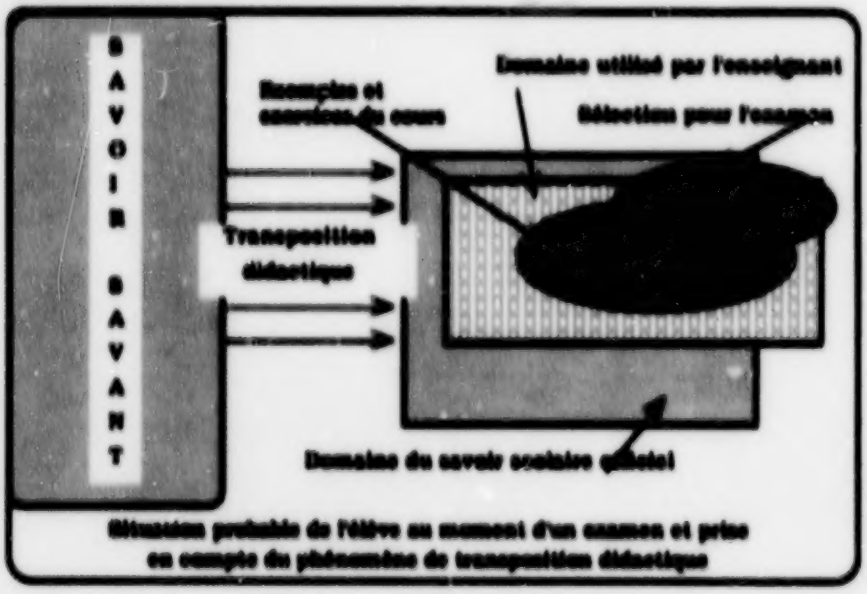

SAVOIR SAVANT
Domaine utilisé par l'enseignant
Transposition didactique
Domaine du savoir scolaire effectif
Situation probable de l'élève au moment d'un examen et prise en compte du phénomène de transposition didactique

Figure 7

De même que se posent des questions de validité, il est à vérifier si telle question d'évaluation peut fournir l'information que l'on souhaite obtenir sur le savoir d'un étudiant, d'un candidat, d'une personne?

Les réflexions qui précèdent m'ont conduit à mettre en parallèle deux façons de communiquer les objectifs: l'une qui s'appuie sur l'anticipation et l'autre, sur la négociation explicite des objets de savoir accompagnés de leurs significations propres (voir *annexe I*).

Nous nous sommes arrêtés peut-être trop longtemps sur cette question des relations entre communication et évaluation et nous nous sommes autorisés bien des détours. Il nous semble que l'enjeu est important. En laissant des méthodologies non contrôlées par la didactique (ce qui ne signifie pas nécessairement contrôlées par les didacticiens) occuper la place, nous risquons en effet de voir s'éloigner le moment où l'enseignement des mathématiques pourra communiquer le pourquoi avec le comment, le sens avec l'opérationnalité. Il semble donc urgent que la recherche en didactique des mathématiques s'intéresse de près aux faits d'évaluation.

Quoi qu'il en soit, remarquons que, au fil des pages précédentes, la plupart des concepts clés de la didactique se sont trouvés comme convoqués: **régulation, contrat didactique, validation, institutionnalisation, transposition didactique**. Qu'on se rassure, ceux qui manquent à l'appel ne tarderont d'ailleurs pas à apparaître!

Peut-être faut-il invoquer ici l'un des principes de la régulation cybernétique qui postule qu'un système régulé est isomorphe à son système de régulation (principe d'Ashby)? Remarquons en passant qu'isomorphie ne signifie pas absence d'éléments communs.

Que faut-il conclure?

Que le système didactique est déjà assez complexe à étudier et que la (science) didactique est encore trop jeune pour se risquer à se disperser en prenant en compte de façon explicite les faits d'évaluation? Que le discours sur l'évaluation ne fait que doubler de façon pâle, inutile, et peut être nuisible l'effort de théorisation qui s'effectue dans le champ de la didactique des mathématiques? Que l'importance pour le fonctionnement du système didactique de l'ensemble des faits d'évaluation est ici qu'il n'est plus possible d'éviter l'affrontement frontal, au moins en certains endroits de l'étude de ce système.

Personnellement, il me semble que c'est la dernière solution qui s'impose.

L'ÉVALUATION ENTRE MESURE ET SENS

Si la notion d'évaluation véhicule avec elle l'idée de **valeur**, au moins par son étymologie, elle véhicule aussi celle d'**incertitude** par son histoire. La disparition de l'incertitude aurait à substituer la notion de **mesure** à celle d'évaluation. L'un des apports des recherches en didactique est sans doute d'avoir mis en évidence l'impossibilité de cette disparition: **le savoir mathématique d'un élève ou d'un groupe d'élèves n'est pas mesurable.**

Dans le cas de l'évaluation scolaire classique, qu'il s'agisse de noter leurs propres élèves ou d'autres élèves (en situation d'examen par exemple), il est pourtant demandé aux enseignants de repérer, sur une **échelle**, la valeur du savoir de ces élèves par rapport à un domaine plus ou moins bien défini. Chaque enseignant sait que cette échelle est personnelle (même lorsque des barèmes communs sont utilisés), souvent non explicitable, variable dans le temps, difficile à mettre en relation avec des significations d'ordre didactique, qu'elle ne présente que peu de garantie de **validité**, de **fidélité**, de **sensibilité**, de **précision**,...

Dans le cas où il s'agit d'évaluer le fonctionnement du système (évaluations nationales, évaluations et enquêtes internationales), le souci de comparaison des professeurs entre eux, des établissements entre eux, ou des pays entre eux, conduit aussi à construire des échelles.

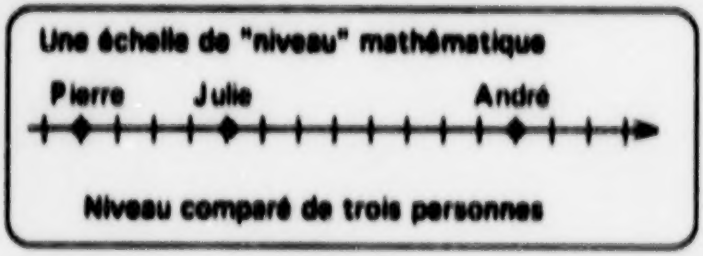


Figure 8

Ces échelles seraient intéressantes si elle concernaient un **champ conceptuel** (ou une classe de questions ayant une cohérence suffisante) et dans lequel on aurait préalablement repéré des hiérarchies. Malheureusement, le souci principal concerne habituellement la fidélité, ce qui finit par être obtenu grâce à une certaine insignifiance. La chose s'aggrave lorsque l'échelle utilisée s'applique au domaine mathématique dans son ensemble. On objectera sans doute qu'il ne s'agit plus là de mesure mais **d'indicateurs** et que tout le monde en est conscient! L'ennui est que ces indicateurs sont utilisés comme des mesures.

La *figure 8* se passe de commentaires! Pour ce qui est de la figure 9, comment passer de là à un classement absolu? Cela ne peut être fait, si cela est nécessaire, qu'au prix d'une réduction que l'on connaît bien. Mais cette réduction suppose elle-même des choix de pondérations qui ne peuvent pas être faits sur des critères psychométriques.

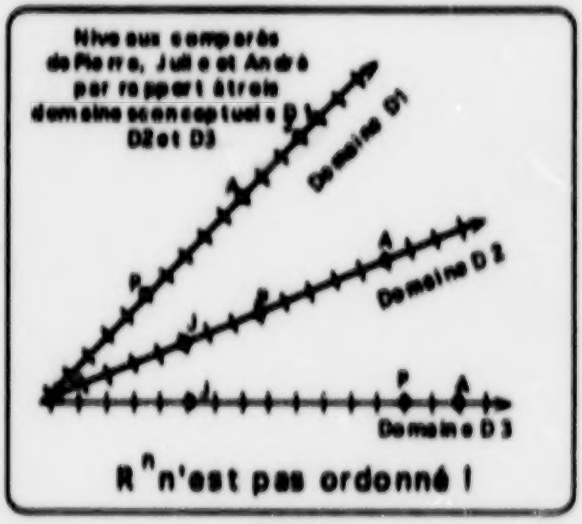


Figure 9

Nous avons pu constater, par exemple, la faible corrélation existant entre les «niveaux» obtenus sur des échelles correspondant à des champs aussi grossiers que «géométrie» d'un côté et «domaine numérique» de l'autre, ou encore entre les «niveaux» mesurés sur des tâches proches de l'apprentissage (niveaux taxinomiques inférieurs) des niveaux mesurés sur des tâches correspondant aux niveaux taxinomiques supérieurs. Là aussi, la psychométrie ne donne pas de clés pour faire les choix nécessaires à une réduction unidimensionnelle.

Un type d'échelle intéressant doit pouvoir être obtenu en utilisant un modèle didactique du fonctionnement des connaissances combiné avec des analyses de données recueillies dans des études à grande échelle. Certaines recherches sont suffisamment avancées pour qu'il soit possible d'envisager de telles échelles. On peut par exemple penser à des domaines tels que celui des situations additives, la proportionnalité, celui de la géométrie de construction, des probabilités.

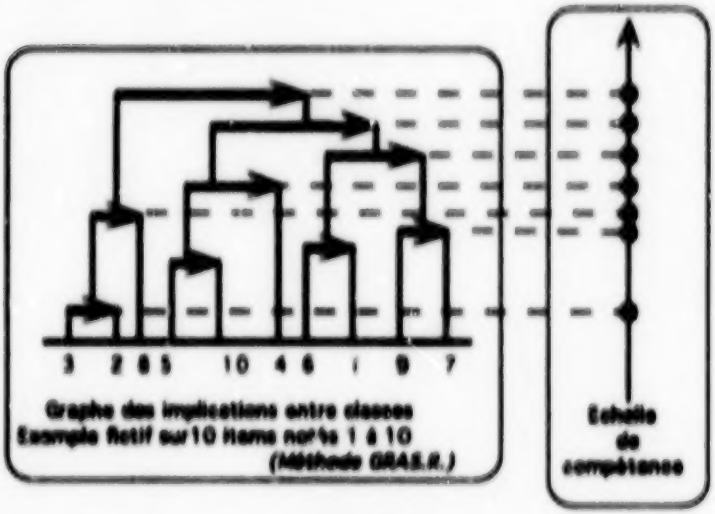

3 2 8 5 10 4 6 1 9 7
Graphe des implications entre classes
Exemple fictif sur 10 items notés 1 à 10
(Méthode GRAS.R.)
Echelle de compétence

Figure 10

Des recherches de ce type sont en cours; elles s'appuient en particulier sur la **théorie des champs conceptuels** (Vergnaud, 1990) et sur une méthode d'analyse originale développée dans le cadre des recherches en didactique: l'**Analyse implicative** (Gras, 1977-1992).

La place ne permet pas de développer cette idée et de plus, les résultats obtenus restent à confirmer. Il est toutefois certain que l'Analyse implicative nous a déjà permis d'obtenir des hiérarchies suffisamment fortes dans plusieurs situations pour pouvoir envisager la construction d'échelles: géométrie de l'espace en Cinquième, calcul vectoriel en Seconde, proportionnalité en Troisième...).

Le schéma de la figure 10 n'est cependant qu'une simplification un peu rapide: la projection seule ne suffira pas à obtenir de bonnes échelles.

Un piste complémentaire intéressante semble être celle de la Théorie de Réponse aux Items (Item Response Theory - Hambleton, R.K.; Swaminathan, H. (1985)). Le lecteur trouvera dans la prochaine partie des courbes caractéristiques d'items qui donnent une petite idée de ce qu'il est possible de proposer. Disons simplement que la méthode utilise les probabilités bayésiennes et les estimateurs statistiques du type «maximum de vraisemblance».

Un travail considérable reste à faire et de plus, il ne faut cependant pas attendre des recherches en didactique qu'elles produisent de telles échelles, directement et de façon immédiatement opératoire. Les recherches sont toujours faites dans un contexte donné et la généralisation ne peut pas être effectuée sans précaution; une généralisation acceptable à un moment donné peut très bien devenir incorrecte avec le temps. Les programmes et les pratiques évoluent et la diffusion d'instruments du type de ceux évoqués n'est pas sans avoir une incidence sur ces évolutions. De telles échelles devraient cependant être élaborées et plus ou moins validées dans le cadre de recherches d'ingénierie didactique, mais on reste, là encore au niveau de la recherche et la dissémination ne peut être faite que par les enseignants, ni sous leur responsabilité.

Notons que la critique de la mesure classique, qui est implicite dans ce qui précède, est aussi le fait des psychométriciens eux-mêmes et que ceux-ci cherchent à mettre au point de nouveaux modèles et de nouveaux instruments qu'il n'y a pas lieu de rejeter a priori.

Pour nous, il s'agirait plutôt de chercher un meilleur équilibre et une meilleure collaboration conduisant à mettre les modèles et les méthodes liés à la mesure sous le contrôle de la signification, donc sous celui du didactique.

À titre d'exemple, voici un avis autorisé du côté de la psychométrie. Des spécialistes de d'Educational Testing Service (Princeton) écrivent (Frederiksen, N., Mislevy, R. J., Bejar, I. I., 1993):

> «*Ce qui manque dans la conception sur laquelle la théorie standard des test est basée sont des modèles pour expliquer comment les gens arrivent à savoir ce qu'ils savent et à faire ce qu'ils savent faire et la façon dont ils améliorent leurs capacités. L'impulsion pour un examen des fondations la théorie des tests vient de psychologues et d'éducateurs, qui, alors qu'ils repoussent les frontières de leur compréhension du domaine cognitif, posent des questions qui sortent de l'univers du discours produit par la théorie classique des tests.*»

Et un autre qui nous semble complémentaire (Wittrock, M. C., 1991):

> «*Le nouveau type de tests fournirait une information de nature diagnostique, relativement aux conceptions des étudiants, à leurs stratégies d'apprentissage, ainsi qu'aux processus de pensée métacognitive et affective impliqués dans la construction du sens...*
>
> *Ces tests des processus cognitifs fourniraient une information pertinente en ce qui concerne les diagnostics d'apprentissage des étudiants ainsi que l'amélioration de l'enseignement. Tout cela à partir d'une meilleure compréhension de la façon dont les élèves apprennent.*»

Ce sont les psychométriciens eux-mêmes qui cherchent à échapper aux conceptions classiques de la mesure et convoitent de nouvelles voies, soit de nouveaux types d'échelles, soit des instruments d'analyse qualitative qui nous ramène à une problématique du sens.

Mais au fait, pourquoi avoir tant insisté sur ces questions d'échelles?

D'une part, de nombreuses échelles implicites ou explicites, sont utilisées partout où il est question d'évaluation: construites avec précaution ou construites n'importe comment (un point n'est-il pas toujours la moitié de deux point?).

D'autre part, la nécessité sociale de classement ou simplement de contrôle, ou encore le souci d'égalité devant les procédures certificatives conduit à souhaiter l'introduction d'échelles.

Notons encore que la tradition académique en France reste très éloignée de la vague des tests standardisés et des préoccupations d'ordre psychométrique qui se sont imposés dans la plupart des autres pays. Il n'est cependant pas certain que la qualité de nos évaluations et en particulier des évaluations certificatives, n'aient pas quelque chose à gagner à une prise en compte prudente des techniques développées ailleurs. Par contre, il est certain qu'elles auraient beaucoup à perdre à une imposition administrative qui ne serait pas précédée de recherches sérieuses.

Pour éviter une certaine ambiguïté, rappelons aussi que cet exposé n'est pas un plaidoyer pour l'évaluation. Mon sentiment, partagé par beaucoup de chercheurs qui, dans le domaine de sciences de l'éducation, s'intéressent aux questions d'évaluation est contraire à l'opinion qu'en France, au moins, il y a déjà trop d'évaluation.

En est-ce fini avec la mesure? Pas tout à fait! En effet nous allons passer maintenant du côté du sens, mais on a déjà pu entrevoir que cela n'était pas nécessairement contradictoire avec l'idée d'échelle. Bien plus, une condition nécessaire à la possibilité de construire de bonnes échelles est la maîtrise du sens de l'information produite par les situations d'évaluation.

Les situations d'évaluations ne sont que déclencheurs de comportements et dans le meilleur des cas, des révélateurs de possible. Il n'y a pas transparence et l'évaluation ne donne pas un accès direct au savoir du sujet: ce qui est observé à propos du savoir d'un élève, aurait pu ne pas être observé, si la situation avait été légèrement modifiée, tout autant que ce qui n'a pas été observé aurait pu l'être.

Ce qui importe dans une évaluation, c'est moins le comportement observé que les inférences que les observations permettent de faire. Une des qualités attendues d'une situation d'évaluation est sa **généralisabilité**, c'est-à-dire ce qu'elle permet de dire sur ce qui n'a pas été directement observé. Ce point est à relier à la question de la définition, de la délimitation et de l'organisation des **compétences**.

Selon la fonction attribuée à l'évaluation, les qualités classiques de la psychométrie gardent leur importance (en particulier la **fidélité** et la **validité**), mais ces qualités devraient être assujetties à la généralisabilité. Notons que, comme cela est souvent le cas, ces deux qualités sont antagonistes (améliorer l'une conduit souvent à diminuer les autres). Il n'est donc pas possible de construire des évaluations qui soient utilisables sans précaution, quelle que soit la fonction attribuée à l'évaluation (formative, diagnostique, sommative, certificative...). Un même instrument d'évaluation peut cependant être utilisé pour des fonctions différentes, à condition que l'interprétation faite des observations soit asservie à cette fonction. De plus, ce qui précède n'est pas contradictoire avec l'idée, exprimée plus haut, du caractère intrinsèquement intéressant des questions d'évaluation.

Le chapitre suivant nous donnera de nombreuses possibilités d'illustrer et d'illustrer la question du sens sous divers aspects.

ANALYSE DE QUELQUES QUESTIONS D'ÉVALUATION

Exemple 1: La bissectrice

Parmi les nombreux exemples que nous pourrions donner, en voici un qui illustre à la fois l'importance du **contexte** dans lequel une question est posée, l'importance des **conceptions** des élèves, l'existence de **schèmes** qui peuvent ou non être activés selon la situation. Cette question illustre aussi implicitement le parti que l'on peut tirer en formation, que de la connaissance du comportement des questions (c'est bien le comportement des questions qui nous intéresse et non plus seulement celui des élèves).

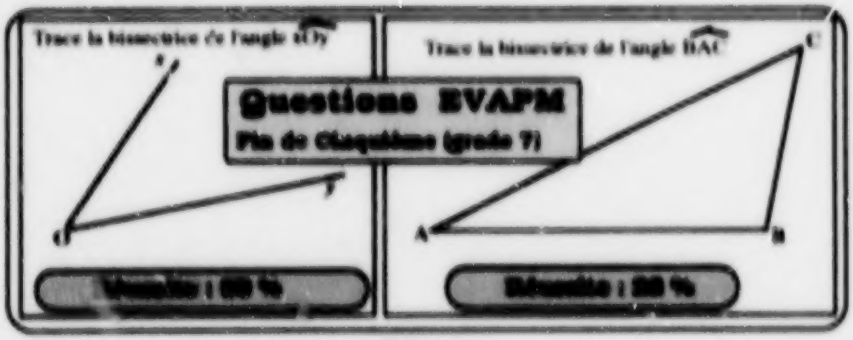


Figure 11

En Sixième (grade 6), la différence importante entre la réussite concernant le tracé de la bissectrice d'un angle selon que cet angle est isolé (70%) où qu'il appartient à un triangle (28%) ne peut manquer d'attirer l'attention.

Dans les deux cas, les élèves utilisent un intermédiaire qui est un milieu, mais dans le second cas ils *préfèrent* prendre le milieu du côté opposé, ce qui en particulier leur permet d'utiliser une règle graduée plutôt que le compas qui leur paraît moins sûr. Insistons sur le fait que pour les élèves les deux procédures sont équivalentes. Nous avons sans doute là un bel exemple de conception «erronée» mais qui a son domaine de validité propre (triangle isocèle) et qui, de plus reste souvent inaperçue. En effet, pour que la différence médiane-bissectrice apparaisse nettement sur la figure, il faut que l'angle B soit assez grand par rapport à l'angle C, ce qui est rarement le cas dans les situations rencontrées par les élèves.

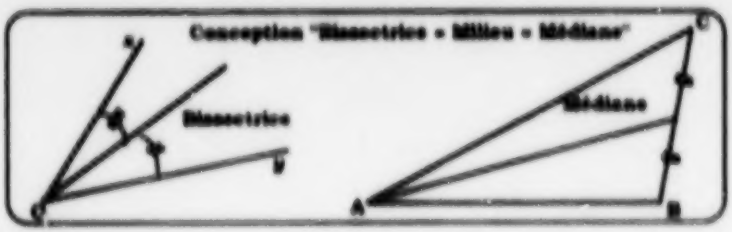


Figure 12

Exemple 2: Angles et triangles

Considérons maintenant les deux questions suivantes, utilisées dans des évaluations à grande échelle (EVAPM), ce qui permet d'avoir des intervalles de confiance de très faible amplitude et de pouvoir construire des courbes de réponses assez fiables (voir plus loin).

La question A7 est un **bon indicateur du niveau de compétence** des élèves du moins en géométrie, tel qu'il peut être habituellement estimé aussi bien par les évaluations des enseignants que par une évaluation indépendante conforme au curriculum. Ce qui signifie en particulier que la question est discriminante. Pour le reste, il faudrait se référer aux explications qui accompagnent les courbes de réponses.

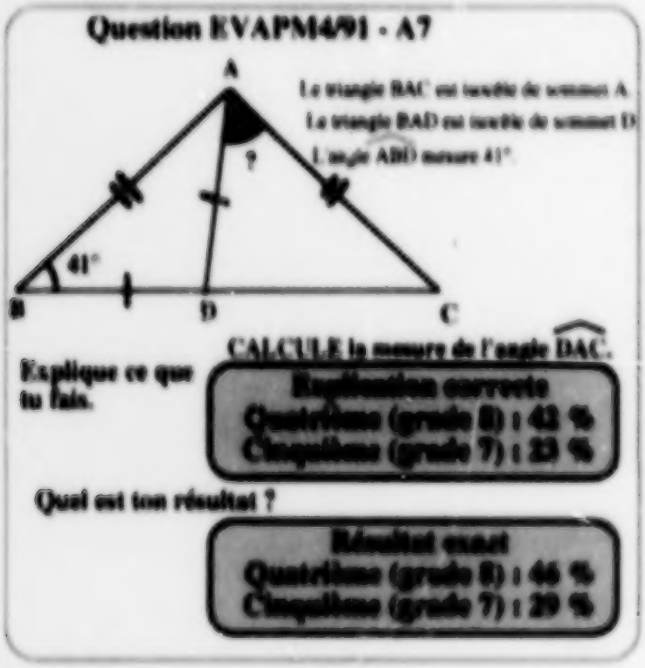


Figure 18

Dans les mêmes conditions, la question A17 se comporte curieusement: la probabilité pour qu'un élève donné réponde correctement à cette question n'est pas une fonction croissante de sa compétence en géométrie, estimée comme ci-dessus (voir encore les courbes de réponses).

L'explication n'est pas bien difficile à trouver mais suppose une observation attentive des procédures utilisées par les élèves. Les élèves «les plus faibles», selon l'indicateur global, manifestent une approche perceptive de la question; ils tentent éventuellement un dessin et **voient** que, selon les cas, la construction est ou n'est pas possible. Ils trouvent ensuite une explication *ad hoc* tout à fait satisfaisante.

Les élèves «les plus forts», selon le même indicateur ont tendance à utiliser directement une procédure déductive, procédure qu'ils maîtrisent assez souvent. Entre les deux, on trouve les élèves qui hésitent: ceux qui seraient plus perceptifs, pas encore déductifs, mais qui chercheraient cependant à appliquer une procédure déductive (théorème), apprise mais non encore maîtrisée.

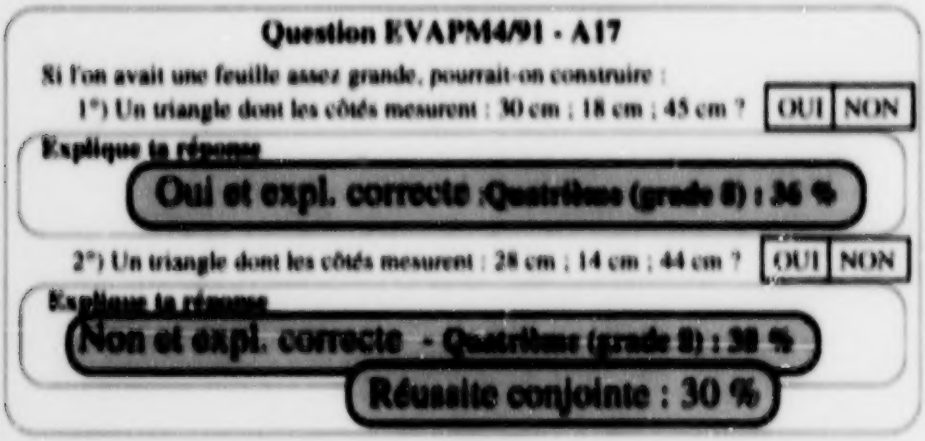

Question EVAPM4/91 - A17
Si l'on avait une feuille assez grande, pourrait-on construire :
1°) Un triangle dont les côtés mesurent : 30 cm ; 18 cm ; 45 cm ?
OUI
NON
Explique ta réponse
Oui et expl. correcte :Quatrième (grade 8) : 36 %
2°) Un triangle dont les côtés mesurent : 28 cm ; 14 cm ; 44 cm ?
OUI
NON
Explique ta réponse
Non et expl. correcte - Quatrième (grade 8) : 38 %
Réussite conjointe : 30 %

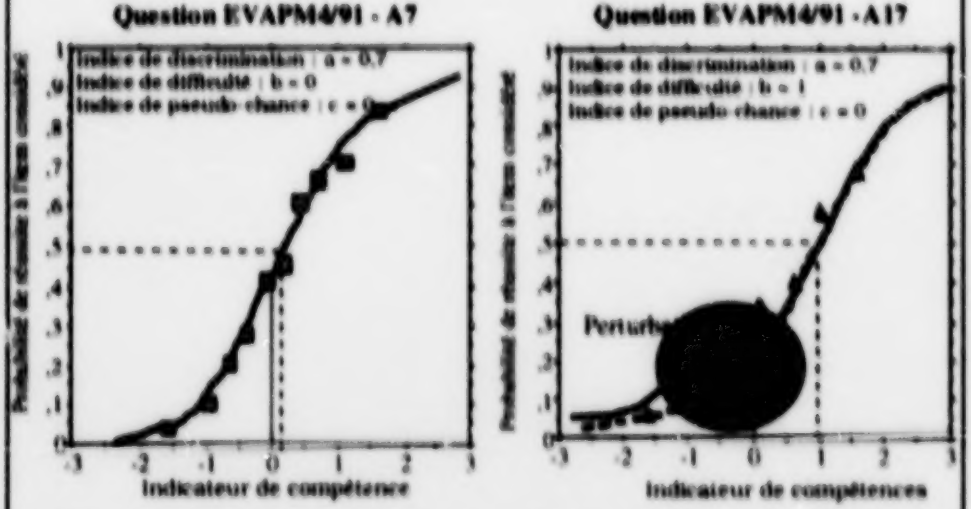
Figure 14

Les courbes de réponses ci-dessous illustrent bien le phénomène observé.

La question A7 peut légitimement être intégrée pour la constitution d'une échelle.

La question A17 ne devrait pas être utilisée pour une telle échelle. En effet, quelle que soit le mode d'élaboration de l'échelle, elle va pénaliser des élèves qui échouent mais qui en fait, sont plus compétents que certains de ceux qui réussissent. Par contre on voit bien l'intérêt que l'on peut tirer de cette question en situation d'**évaluation diagnostique** strictement qualitative.

Courbes de réponses des questions A7 et A17
(courbes de RASH)


Question EVAPM4/91 - A7
Indice de discrimination : a = 0,7
Indice de difficulté : b = 0
Indice de pseudo-chance : c = 0
Indicateur de compétence
Question EVAPM4/91 - A17
Indice de discrimination : a = 0,7
Indice de difficulté : b = 1
Indice de pseudo-chance : c = 0
Indicateur de compétences

Un indicateur en abscisse est placé de compétence; celui-ci est normalisé afin de permettre les comparaisons.

Chaque point du graphique (petits triangles ou petits carrés selon le cas) correspond à 10% de la population (décalage basé sur la valeur de l'indicateur utilisé en abscisse). Par exemple, le point le plus à droite représente les 10% des élèves qui réussissent le mieux selon cet indicateur.

La taille importante des échantillons traités assure une grande fiabilité aux courbes obtenues. Le remplacement d'un indicateur de référence par un autre qui lui serait fortement lié ne modifie que très peu les courbes obtenues (robustesse du procédé).

Par définition, l'indice de discrimination est proportionnel à la pente de la tangente à la courbe au point d'ordonnée 0,5.

Les deux courbes présentées ici concernent des items ou questions de difficultés nettement différentes mais ayant un pouvoir discriminatif égal et relativement important.

D'un point de vue psychométrique, la question A7 a un bon comportement; mais en ce qui concerne l'analyse didactique, c'est la question A17 qui présente le plus d'intérêt.

Figure 18

Exemple 3: Les alignements

Les méthodes relatives à la Théorie de Réponse aux Items (IRT déjà évoquée) conduisent à construire des échelles de compétences indépendantes des groupes d'élèves soumis aux évaluations. Cela règle en théorie la question de la **difficulté** des questions, difficulté qui, dans la conception classique, ne pouvait qu'être relative à un groupe donné.

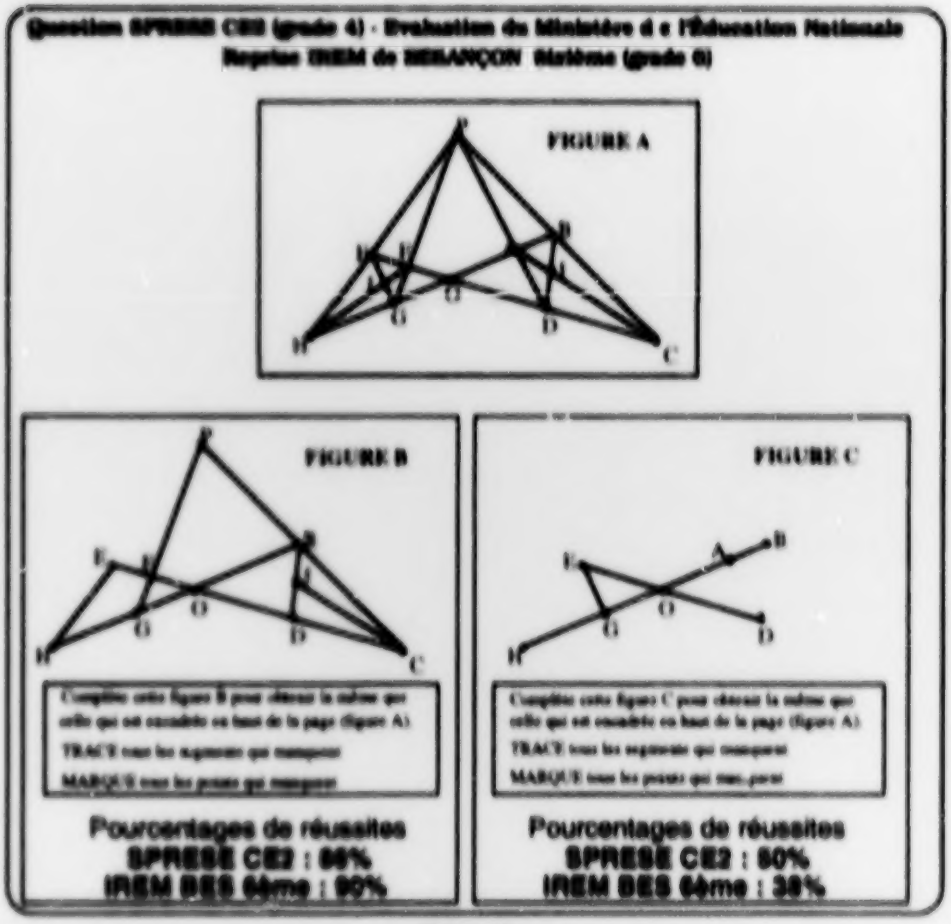

Question SPRESE CE2 (grade 4) - Evaluation du Ministère de l'Education Nationale
Reprise IREM de BESANÇON Sixième (grade 6)
FIGURE A
FIGURE B
FIGURE C
Pourcentages de réussites
SPRESE CE2 : 66%
IREM BES 6ème : 90%
Pourcentages de réussites
SPRESE CE2 : 60%
IREM BES 6ème : 38%

Figure 18

Il est certain qu'une notion de difficulté intrinsèque (ou de facilité) serait la bienvenue, surtout si elle n'entre pas en contradiction avec d'autres notions importantes.

La question présentée figure 16 a été utilisée en fin et au CE2 (grades 4 et 6). Il se trouve que cette question est mieux réussie au CE2 qu'en Sixième.

Bien entendu on pense **contrat didactique**..., mais on peut être plus précis.

Signalons d'abord que, dans le cadre du Suivi Scientifique des programmes 1987, cette question a été un bon révélateur des difficultés de certains élèves. Aucun élément de vocabulaire ni aucune autre connaissance formelle n'étant en jeu, les difficultés concernent, selon les élèves, la capacité d'explorer une figure et d'en appréhender les éléments et les relations constitutifs, ou la capacité à coordonner la manipulation des instruments de dessin.

Dans la figure C, le point P doit être placé en premier, à moins de mesurer et de prendre en compte les propriétés métriques de la figure (au lieu des seules propriétés topologiques). C'est ce que font certains élèves de Sixième.

Au CE2, les élèves utilisent une procédure affine et ne prennent en compte que les alignements. En Sixième, où l'on utilise plus les instruments de mesure qu'au CE2 (règle graduée), les élèves ont tendance à essayer de s'en servir. L'aspect métrique de la figure est beaucoup plus souvent pris en compte qu'au CE2. Mais chacun admettra qu'il faut être plus adroit pour réussir avec la procédure métrique qu'avec la procédure affine.

Dans le cas de cette question, l'utilisation d'une échelle de compétences relatives aux tracés géométriques qui couvrirait, disons, l'école élémentaire et le début de l'enseignement secondaire risquerait fort de fournir une courbe de réponse de l'un des types A et B proposés (figure 17).

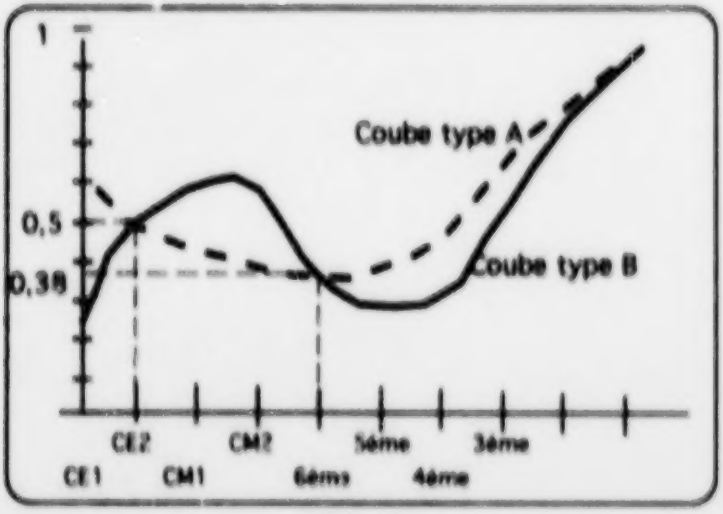


Figure 17

Cette question, qui apporte une information intéressante sur les relations que l'élève entretient avec les objets géométriques, ne devrait pas non plus pouvoir entrer dans la constitution d'une échelle.

Il ne serait pas non plus correct de dire, comme on l'a vu faire dans ce cas, que le niveau des élèves aurait baissé entre le CE2 et la Sixième.

Les questions peuvent rester apparemment les mêmes, mais en fait, elles changent au fur et à mesure que changent les objets de savoir.

Les questions d'évaluation peuvent ainsi se comporter comme des révélateurs de l'état du **contrat didactique**. Elles peuvent aussi mettre en évidence les **conceptions** personnelles des élèves. Enfin, selon la façon dont on les regarde, elles renseignent sur l'état du **rapport au savoir**: le rapport que l'élève entretient avec un savoir particulier, éventuel savoir officiel... (Chevallard, 1989).

Ces trois concepts de la didactique peuvent être utilisés pour rendre compte du fonctionnement d'une question. Ils ne sont pas pour autant équivalents et chacun apporte son éclairage.

Exemple 4: Les pommes

L'un des intérêts des évaluations à grande échelle est qu'elles permettent de suivre l'évolution dans le temps des comportements des élèves par rapport à des questions données.

Grâce à l'observatoire EVAPM, nous disposons de questions suivies de la Sixième à la Première.

Question Observatoire EVAPM

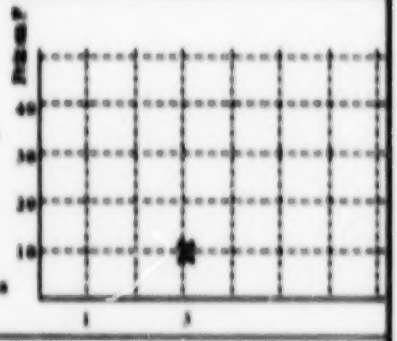
Un producteur de pommes a fixé le prix proportionnellement à la quantité.
Sur un graphique, on a croix indique le prix de 7 kg de pommes.

Indiquez de la même manière (par une croix) :

a) : Le prix de 4 kg de pommes

b) : Le prix de 7 kg de pommes

c) : La quantité de pommes que l'on a pour 15



Pourcentages de réussite

Fin de Cinquième (grade 7) 1988 :	a) 47%	b) 43%	c) 31%
Fin de Quatrième (grade 8) 1989 :	a) 65%	b) 67%	c) 55%

Figure 18

Pour ne pas trop alourdir la présentation, nous nous contenterons ici d'une question posée en fin de Cinquième (grade 7), puis en fin de Quatrième (grade 8) (figure 18).

Au niveau Cinquième, peu d'élèves ont tracé la demi-droite pour placer les croix. La plupart des élèves qui ont réussi cet exercice l'ont fait par le calcul (recherche du prix de 1 kg); l'usage de la calculatrice a favorisé cette stratégie. À ce niveau, la question a donc une validité assez limitée.

Au niveau Quatrième, on observe une amélioration des taux de réussite mais aussi un changement dans les procédures utilisées. Le passage par le calcul devient rare. Les élèves tracent la demi-droite et placent ensuite les croix demandées. Cette évolution est significative d'une acquisition. **Théorème en acte** ou théorème explicite?: *des points de la représentation graphique d'une relation de proportionnalité sont alignés entre eux et alignés avec l'origine des axes*. S'il [illegible] de savoir si les élèves ont une bonne maîtrise de cette

représentation, il est clair que la question présente une validité plus grande au niveau Quatrième qu'au niveau Cinquième.

En Cinquième comme en Quatrième, la réussite est meilleure dans le sens (de l'objet à l'image) que de l'image à l'objet). Cela est assez bien connu et cela est vrai aussi bien dans le cas du passage par le calcul que dans le cas de la lecture sur une représentation graphique. On trouve la trace de cette difficulté dans des questions posées en fin de Seconde (difficulté à utiliser une représentation graphique pour des résolutions approchées d'équations ou d'inéquations).

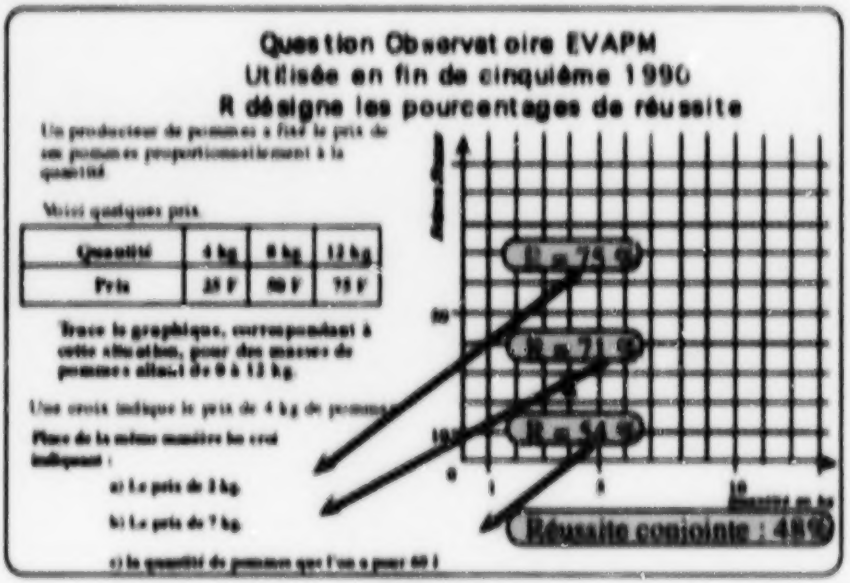

Question Observatoire EVAPM
Utilisée en fin de cinquième 1990
R désigne les pourcentages de réussite
Un producteur de pommes a fixé le prix de ses pommes proportionnellement à la quantité.
Voici quelques prix.
Quantité | 4 kg | 8 kg | 12 kg
Prix | 25 F | 50 F | 75 F
Trace le graphique, correspondant à cette situation, pour des masses de pommes allant de 0 à 12 kg.
Une croix indique le prix de 4 kg de pommes.
Place de la même manière les croix indiquant :
a) Le prix de 2 kg.
b) Le prix de 7 kg.
c) la quantité de pommes que l'on a pour 60 f.
R = 75 %
R = 71 %
R = 54 %
Réussite conjointe : 48 %

Figure 19

La question a été reprise en 1990, en fin de Cinquième, sous une forme modifiée.

Dans cette opérationnalisation, les informations données sont largement redondantes et l'élève est largement aidé par la consigne préalable, indiquant de «tracer le graphique». L'interprétation est faite ensuite avec peu de difficultés, mais on notera que la difficulté de lecture dans le sens (de l'image vers l'objet) subsiste largement.

La modification a été faite à moitié par curiosité, moitié pour rester plus proche «de ce que *l'on savait que les élèves étaient capables de faire*». Donc pas de surprise. Mais la phrase entre guillemets semblera quelque peu assassiné pour l'évaluation. En effet pourquoi continuer à chercher si l'on sait déjà?

En fait, par rapport à la question de la figure 18, la question ci-dessus est un bon exemple de l'effet Topaze (Brousseau, 1986). On fait un peu comme si on gardait la même question, mais on introduit le ses aux moutons/as/ qui permet aux élèves de reconnaître à défaut de connaître.

L'expérience accumulée autour de l'observatoire EVAPM (quelque 2 000 questions) nous permet d'affirmer (à peu près) que l'on est maintenant capable de faire des évaluations sur mesure: «donnez-nous les résultats souhaités et nous vous donnerons les questions».

Exemple 5: Les maisons

Certaines questions ne sont liées à aucune «mesure» de compétence, telles qu'on les établit habituellement: évaluation de l'enseignant, épreuves «objectives», etc.

C'est par exemple le cas de la question suivante:

Question EVAPM

posée en Troisième (grade 9) et en Seconde (grade 10)

Trois personnes, de trois nationalités différentes, habitent les trois premières maisons d'une rue: chaque maison a une couleur différente et chaque personne un métier différent.

A - Le Français habite la maison rouge.

B - L'Allemand est musicien.

C - L'Anglais habite la maison du milieu.

D - La maison rouge est à côté de la verte.

E - L'écrivain habite la première maison à gauche.

Quelle est la nationalité de l'écrivain? et qui habite la maison jaune?

Réponses correctes, avec ou sans explications: **Troisième: 62 % - Seconde: 68 %**

Explications convaincantes: **Troisième: 44 % - Seconde: 38 %**

Figure 20

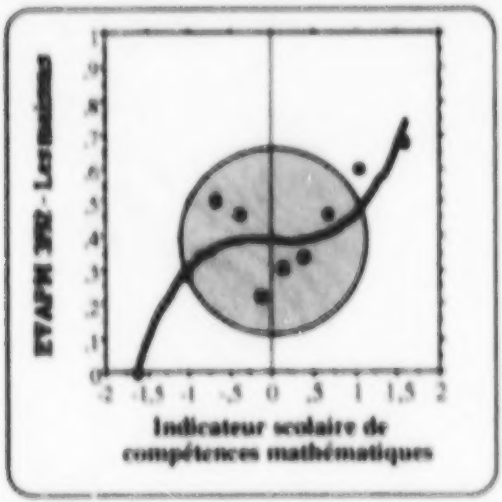


Figure 21

La «courbe» de réponses relative à ce petit problème parle d'elle-même. Il n'est d'ailleurs pas légitime de tracer cette courbe.

D'une façon générale, tous les problèmes qui s'éloignent des pratiques scolaires et en particulier ceux qui supposent la mise en oeuvre de qualités d'imagination, de créativité, sont ainsi peu corrélés avec les indicateurs habituels de compétences scolaires.

D'une façon apparemment contradictoire, il en est de même, au niveau du Collège (grade 6 à 9) pour les questions du domaine statistique, qu'elles soient globalement bien ou mal réussies.

Nous pourrions présenter bien d'autres questions et illustrer d'autres thèmes. Nous renvoyons aux actes de la 45ième rencontre de la CIEAEM pour une analyse d'une épreuve complète utilisant l'analyse implicative (Bodin, 1993).

A propos de la question du sens, nous renvoyons à la «typologie de la signification» (Bodin, 1988). Cette typologie permet de distinguer les questions d'évaluation selon qu'elles font intervenir l'**aspect outil** ou l'**aspect objet** des notions concernées (Douady, 1984); elle permet aussi de distinguer les questions selon qu'elles interpellent plutôt les conceptions ou encore selon qu'elles font intervenir les notions pour résoudre des problèmes de communication (par exemple codages de figures).

Renvoyons finalement à la taxonomie de la complexité cognitive (Gras, 1977) ainsi qu'à la typologie d'activités associées.

De même que l'utilisation des taxonomies de la complexité ont permis de mettre en évidence que l'évaluation s'adressait le plus souvent aux niveaux les plus bas de ces taxonomies, de même l'utilisation d'une typologie de la signification met en évidence le fait que la plupart des questions d'évaluation concernent l'aspect objet des notions en jeu.

L'utilisation de ces typologies, associées à l'utilisation de plans d'évaluation, du type tables de spécification, permettent au moins d'éviter les évaluations caricaturales que l'on voit encore souvent.

CONCLUSION

La qualité de l'évaluation, c'est-à-dire la qualité des informations recueillies et des inférences faites, conditionne dans une large mesure la qualité des didactiques.

Dans le cadre de la recherche en didactique des mathématiques, l'évaluation est un gigantesque chantier à peine entamé.

Des pistes semblent se présenter pour des travaux qui se donneraient comme objectif une meilleure compréhension des processus qui accompagnent les actes d'évaluation.

Des pistes pour des avancées méthodologiques ancrées dans la réflexion didactiques existent aussi.

Il n'y a plus qu'à travailler pour exploiter au mieux ces pistes tout en conservant la prudence qui, dans le domaine habituellement largement idéologique de l'évaluation, s'impose plus qu'ailleurs.

ANNEXE

Deux façons de concevoir la communication des objectifs

	Anticipation	**Négociation**
	Communiquer ses objectifs pour permettre à l'élève d'anticiper l'évaluation.	Communiquer les objectifs pour aider l'élève à donner du sens à ses apprentissages, y compris face à l'évaluation.
A **Information Communication**	*Communication des objectifs de la formation* **Enseignant** ↓ **Apprenant**	*Explicitation progressive et interactive des valeurs, buts et finalités* **Enseignant** ↓ ↑ **Apprenant**
B **Formation**	*Intériorisation par l'apprenant des objectifs de l'enseignant (objectifs que l'apprenant finit, si tout se passe bien, par faire siens).* **Enseignant** ↓ **Apprenant** ↺	*Intégration progressive par l'élève de SAVOIRS ouverts de significations et de valeurs de nature épistémiques (Apprentissage)* **Enseignant** ↓ ↑ **Apprenant** ↺
B' **Préparation aux examens et aux contrôles**	*Anticipation par l'élève des questions d'examen ou de contrôle* **Enseignant** ↷ **Apprenant**	*Analyse de l'écart existant entre les objectifs définis en A et les savoirs résultant de B* **Enseignant** ↺ ↓ ↑ **Apprenant** ↺
C **Examens Contrôles**	*Situations opérationnalisant les objectifs de l'enseignant* **Enseignant** ↻ **Apprenant**	*Mise à plat de l'écart entre les objectifs définis en A et les savoirs résultant de B* **Enseignant** ↻ **Apprenant** ↺
Remarques diverses **Avantages et inconvénients**	*Perte de vue des significations profondes des objets d'enseignement au bénéfice d'un glissement vers les conditions d'évaluation. Perte de vue de la pertinence épistémologique (du sens profond des questions étudiées).* ***Réaliste et efficace** au prix d'un certain conditionnement de l'apprenant.* ***Risque** de renforcer l'écrasement de l'apprentissage sur les objectifs selon une formule de Guy BROUSSEAU.*	*Cohérence épistémologique (respect de la signification des objets d'enseignement)* *Favorise l'autonomie de l'apprenant.* *Permet à l'évaluation de mieux remplir son rôle régulateur (selon Y. CHEVALLARD, l'évaluation a tendance à perdre cette fonction de dire quelque chose de vrai sur le savoir d'autrui).* ***Adapté** à une conception constructiviste de l'apprentissage.*

BIBLIOGRAPHIE

Allal, L., Cardinet, J., Perrenoud, P. (1978). L'évaluation formative dans un enseignement différencié. *Actes du colloque de Genève*, mars, Éditions Peter Lang.

Allal, L., Bain, D., Perrenoud, P. (1993). *Évaluation formative et didactique du français*. Delachaux et Niestlé.

Bain, D. (1988). L'évaluation formative fait fausse route: de la nécessité de changer de cap. *Mesure et évaluation en éducation, 10(2)*.

Bodin, A., et al. *Brochures EVAPM*, contenant les épreuves, les résultats et les analyses des évaluations des programmes de mathématiques menées par l'APMEP (Association des professeurs de mathématiques de l'enseignement public).

Bodin, A. (1988). Evaluation in mathematics: the quality of students' knowledge. *6e congrès international sur l'enseignement des mathématiques (ICME 6)*. Budapest, août. Traduction française: l'évaluation du savoir mathématique. *Bulletin de l'Association des professeurs de mathématiques de l'enseignement public (APMEP)*, N°368/1989, pp. 195-219.

Bodin, A. (1992). Évaluation et didactique - distinguer ou intégrer, dans: D. Lavault (dir.): *Les pratiques d'évaluation en éducation*. Montréal: Éditions de l'ADMEE. XVe Journées d'étude de l'ADMEE - Canada, Ottawa-Hull, octobre.

Bodin, A. (1992). Réflexions sur les représentations, les conceptions et les compétences. Recherches en didactiques: contribution à la formation des maîtres. *Actes du colloque INRP*, février, et *Petit x*.

Bodin, A. (1992). What does to assess mean. Conférence ICMI sur le thème «Assessment in mathematics education and its effects», Colonge (SPAIN) Avril 1991 et ICME 7-Québec. in: *Investigations into Assessment in Mathematics Education* - An ICMI Study (ed Mogens NISS) - Kluwer academic publisher, 1993.

Bodin, A. (1993). Développement d'une base de données d'évaluation en mathématiques: EVAPMIB (avec F. Couturier). *Pédagogies N°9/1993*, (Cahiers du Laboratoire de pédagogie expérimentale de l'Université de Louvain).

Bodin, A. (1994). L'évaluation en mathématiques - continuités et perspectives. *Actes de la 45e rencontre CIEAEM*, Cagliari, Italie, juillet 1993.

Brousseau, G. (1986). Fondements et méthodes de la didactique des mathématiques. *Recherches en didactique des mathématiques, 7(2).* Grenoble: Éditions La pensée sauvage.

Brousseau, G. (1979). Évaluation et théories de l'apprentissage en situations scolaires (Polycopié communiqué).

Cardinet, J. (1986). *La maîtrise, communication réussie.* Institut Romand de documentation pédagogique.

Chevallard, Y. (1985). *La transposition didactique - du savoir savant au savoir enseigné.* Grenoble: Éditions La pensée sauvage.

Chevallard, Y. (1989). Évaluation, véridiction, objectivation - la relation didactique comme caprice et miniature, in: *L'évaluateur en révolution.* Actes du colloque ADMEE 89, INRP.

Chevallard, Y., Feldmann, S. (1986). *Pour une analyse didactique de l'évaluation.* IREM d'Aix Marseille.

Douady, R. (1984). *Jeux de cadres et dialectique outil-objet.* Thèse de doctorat, Université de Paris VII.

Fournier, C. (1991). *Étude corrélationnelle des liens entre l'anticipation, la préparation, auto-évaluation et le résultat à un examen de rendement scolaire.* Thèse de doctorat, Faculté d'éducation, Université d'Ottawa.

Frederiksen, N., Mislevy, R.J., Bejar, I. (Eds) (1993). *Test Theory for a new generation of tests.* Educational testing service, Lawrence Erlbaum Associates.

Galperine, P.I. (1966). Essai sur la formation par étapes des actions et des concepts, in: *De l'enseignement programmé à la programmation des connaissances.* Presses Universitaires de Lille, 1980.

Genthon, M. (1991). Communiquer, quoi, pour quoi faire et comment, dans un processus d'évaluation? in: *L'évaluation: problème de communication.* Delval, Suisse.

Gras, R. (1977). *Contributions à l'étude expérimentale et à l'analyse de certaines acquisitions cognitives et de certains objectifs didactiques en mathématiques.* Thèse, Université de Rennes, (description et utilisation de la taxonomie).

Gras, R. (1986). Recherches sur l'apprentissage: Analyse des correspondances et méthodes statistiques apparentées. *Cahier du Crude,* Université du Québec à Montréal.

Gras, R. (1992). L'analyse des données: une méthodologie de traitement de questions de didactique. *Recherches en didactique des mathématiques, 12(1)*. Grenoble: Éditions La pensée sauvage.

Hambleton, R.K., Swaminathan, H. (1985). *Item response theory. Principles and applications*. Kluwer Nghoff Publishing.

Johsua, S., Dupin, J.J. (1993). *Introduction à la didactique des sciences et des mathématiques*. PUF, Paris.

De Ketele, J.M. (1993). L'évaluation conjuguée en paradigmes. *Revue française de pédagogie, 103*.

Margolinas, C. (1993). *De l'importance du vrai et du faux dans la classe de mathématiques*. Recherches en didactique des mathématiques. Grenoble: Éditions La pensée sauvage.

Perrenoud, P. (1991). Pour une approche pragmatique de l'évaluation formative. *Mesure et évaluation en éducation, 13(4)*.

Vergnaud, G. (1981). *L'enfant, l'école, et la réalité*. Peter Lang.

Vergnaud, G. (1990). La théorie des champs conceptuels. *Recherches en didactique des mathématiques, 10(2,3)*. Grenoble: Éditions La pensée sauvage.

Vygotski, L.S. (1935). Interaction between learning and development. *Mind in society (1978)*. Harvard University Press.

Wittrock, M.C.(1991). Cognition and testing. *In: Testing and cognition*. Prentice-Hall Inc.

Des exemplaires du présent document sont en vente à l'adresse suivante:

Centre interdisciplinaire de recherche
sur l'apprentissage et le développement en éducation
Université du Québec à Montréal
Pavillon Berri
505, boulevard René-Lévesque Est
Local WB-2220

Case postale **8888**, succursale Centre-ville
Montréal (Québec)
H3C 3P8

Tél.: (514) 987-6186

Publié par le Centre interdisciplinaire de recherche
sur l'apprentissage et le développement en éducation
Université du Québec à Montréal
Conception graphique: Pierrette Thérien